IMAGE FUSION AND ENHANCEMENT USING TRIANGULATED IRREGULAR NETWORKS

G. Scarmana
School of Civil Engineering and Surveying
University of Southern Queensland, AUSTRALIA

ABSTRACT

A triangulated irregular network (TIN) is a viable structure for vector representation of raster image data. To visualize the image characterized by triangulation, it is required to fit a continuous surface of pixel brightness values in the triangulation (i.e. to interpolate data stored in its vertices).

From this perspective, this paper presents a multi-frame image fusion and enhancement process that employs TIN structures rather than arrays of pixels as the original working units. The feasibility of this application relates to the fact that a TIN model offers a good quality digital image representation with a reduced density of pixel values as compared to a corresponding raster representation [4].

In the proposed process several low-resolution unregistered and compressed images (such as those extracted from a video footage) of a common scene are: (a) registered to a sub-pixel level (b) transformed to a TIN structure, (c) grouped or mapped globally within a singular framework to create a denser TIN composite, and (d) the TIN representation is used in reverse to reconstruct a higher resolution image in raster format with more details than any of the original input frames.

Tests and subsequent results are shown to demonstrate the validity and accuracy of the proposed multi-frame image enhancement process. A comparison of this process of multi-frame image enhancement using various interpolation methods and practices is included.

Keywords: Image enhancement, Image fusion, TIN models.

INTRODUCTION

As shown in Figure 1, a digital image can be represented as a TIN structure. This is because the pixels of a raster based image can be considered as 3D points in space in which the x and y coordinates are the rows and columns of the image, whereas the z-coordinate is either the grey or the colour brightness value of a particular pixel.

One of the main advantages of a TIN is that it can be used to approximate a raster image to a selected accuracy with less number of pixels which become nodes in the TIN. The number of pixels used to create the TIN is the smallest number that satisfies two conditions. First, the output TIN must cover the entire area of the input image. Second, a user-specified tolerance *d* must be met.

The tolerance *d* may be defined as the maximum allowable difference in pixel units between the pixel brightness values of the input raster image and the pixel brightness value of the output TIN. In other words, *d* is a factor that contributes to the final accuracy with which the original image will be reconstructed from its TIN representation.

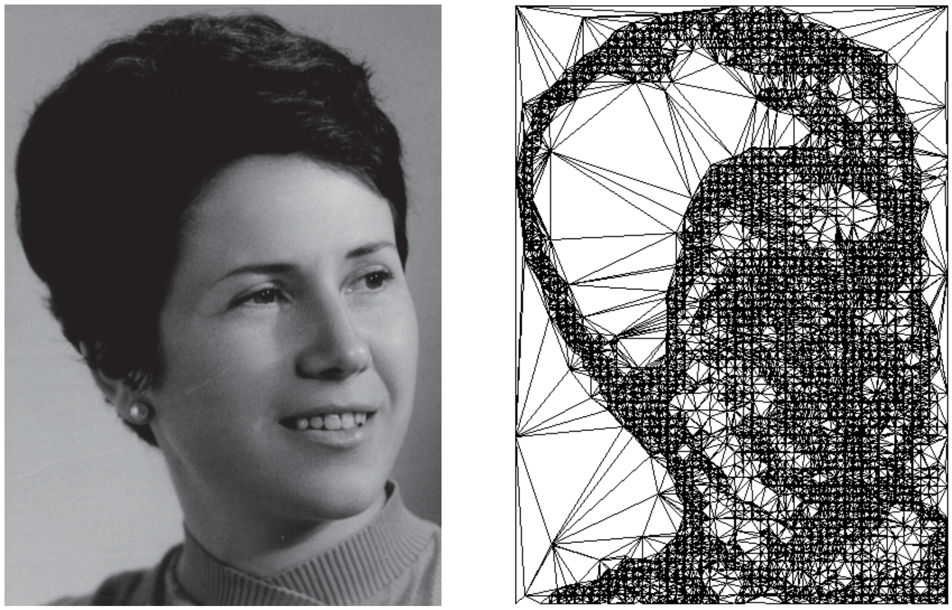
For instance, a large *d* relates to a TIN that conforms less closely to its raster image counterpart. However, the output TIN has fewer pixels (vertices) and triangles and the conversion process to a raster based image is faster. On the other hand, a small *d* makes the TIN conform more closely to a raster based representation. The TIN has more pixels (vertices) and triangles, and will take longer to process. Overall, a TIN representation may often be more compact than a raster based image as it minimizes the amount of redundant data with the subsequent image storage, compression and transmission benefits.

In this context, a TIN image representation is applied here in the particular case of enhancing video images of a common scene. Video sequences usually contain a large overlap between consecutive frames and areas of the scene are sampled in several images at different (and random) sub-pixel positions. This multiple sampling can frequently be combined or fused to achieve images with a higher spatial resolution [6].

Differently from raster multi-frame image enhancement methods, the proposed approach considers TIN structures as the primitive working unit. Figure 2 shows the TIN structures of two slightly shifted images of luisa in Figure 1(a).

Videometrics, Range Imaging, and Applications XIV, edited by Fabio Remondino, Mark R. Shortis, Proc. of SPIE Vol. 10332, 103320V · © 2017 SPIE
CCC code: 0277-786X/17/$18 · doi: 10.1117/12.2279443

Although the two images characterize the same scene they display a number of differences. For the same TIN tolerance *d*, triangles in one image have vanished while new and different triangles have emerged in the other.



(a) (b)

Figure 1. A grey-scale image of luisa.tif (a), and the corresponding triangulated version (b). *d* in this case was equal to 5.

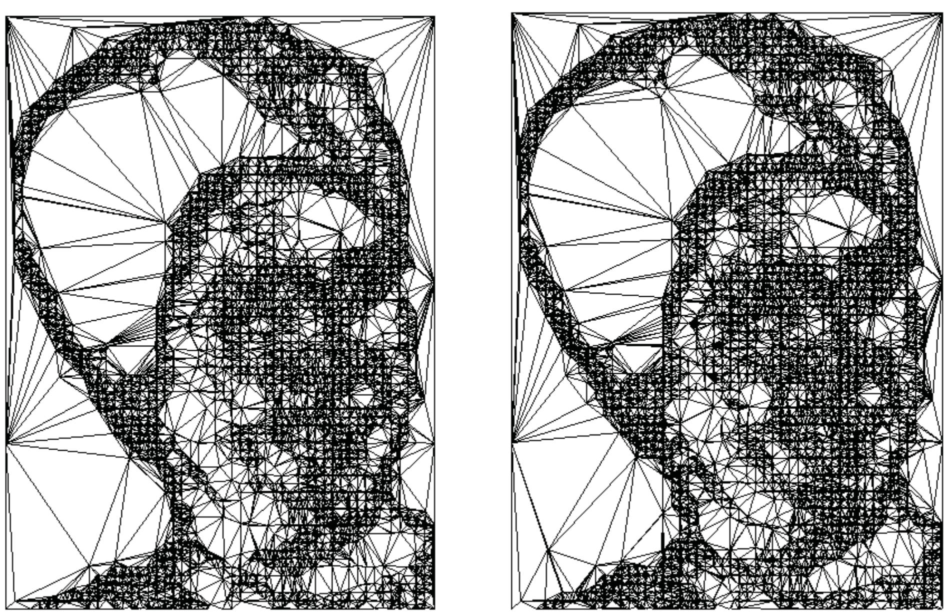


Figure 2. Two TIN representations of the same luisa.tif at different sub-pixel positions. For the same tolerance *d*, the differences between the two are perceivable.

Hence, a composite image can be assembled by fusing the triangulated representations of a sequence of shifted images of the same object within a single higher-resolution reference framework. The result of this process will include more triangle and vertices (i.e. more pixels) than any of the initial input frames. Once a composite or combined triangulated "map" is attained a reverse process is applied so as to transform this TIN composite back to a raster format.

In summary, the proposed image enhancement method involves the following steps:

1. *Sub-pixels image matching*: Estimation of sub-pixel shifts among the different low-resolution input images of a common scene.
2. *TIN representation*: Each low-resolution frame of the sequence is converted to a TIN structure in accordance with a selected tolerance *d*.
3. *Triangulation grouping*: The vertices of the triangles extracted from the input images are merged and/or mapped within a higher resolution common reference frame using the shifts computed in step 1.
4. *Image reconstruction*: An enhanced image in raster format for final display, analysis and use is reconstructed by way of interpolation techniques.

SUB-PIXEL IMAGE MATCHING

There exists a number of methods for determining sub-pixel translation parameters (i.e. in x and y) existing between two shifted images depicting the same scene. The method used here was devised in [2] and is based on DCT (Discrete Fourier Transforms) and normalized cross-correlation registration techniques.

The method allows images to be registered or matched without using control points in the matching procedure. The accuracy of the registration is user selectable and can achieve accuracies of up to 0.01 of a pixel. For two given images (i.e. M and N) the registration outputs the normalized root- mean-squared error (NRMSE.) between M and N, their global phase angle, and the row (x) and column (y) sub-pixel shifts between the two images respectively.

The use of this sub-pixel translation estimator for reconstructing a higher resolution image from a set of low-resolution images is feasible if the input low-resolution images have been under-sampled and therefore contain aliasing. Because of this aliasing, the high-frequency content of the desired reconstruction image is embedded in the low-frequency content of each of the observed images [8].

Hence, given a sufficient number of observation images, and if the set of observations vary in their phase (i.e. if the images of the scene are shifted by a sub-pixel amount), then the phase data can be used to disperse the aliased high-frequency content from the real low-frequency content, and an improved resolution image can be correctly reconstructed [7].

TIN CREATION

A TIN is formed by nodes, triangles and edges. Nodes are locations defined by pixel locations x, y and the pixels brightness values z from which a TIN is constructed. Triangles are formed by connecting each node with its neighbours according to the Delaunay criterion [1]. By using this method the triangles are as equi-angular as possible and any xyz point on the image is as close as possible to a node, and the triangulation results is independent of the order the points are processed. Edges are the sides of the triangles.

A key advantage of the TIN structure is that the density of sampled points, and therefore the size of triangles, can be adjusted to reflect the surface (image) being modelled with more points sampled in image areas of abrupt changes in brightness values.

First, a convex hull is created for a dataset - the smallest convex polygon that contains the set of pixel points. Next, straight lines that do not cross each other are drawn from interior points to points on the boundary of the convex hull and to each other. This distributes the convex bull into a set of polygons which are then divided into triangles by depicting more lines between vertices or the polygons [5].

A TIN is a topological data structure that manages information about nodes comprising each triangle and the neighbors of each triangle. As with other topological data structure, information about a TIN may be conveniently stored in a file or database table, or computed on the fly.

TINs also incorporate the original sample points (pixels) providing a valuable check on the accuracy of the model. Secondly, the variable density of triangles means that a TIN is an efficient way of storing images that have significant pixel brightness variations.

A number of tests by the author have shown that tolerances (or differences) of 2-3 pixel units were sufficient to reconstruct grey-scale images of relative high entropy values. Values of *d* greater than this figure would generate issues of continuity within the reconstructed images.

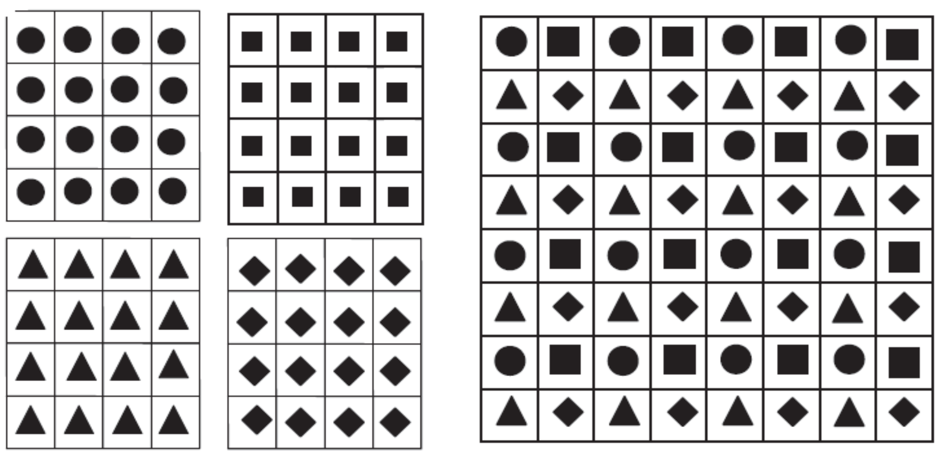
A thorough theoretical explanation of TIN structures is beyond the scope of this work and the reader is referred to [3] for the theory, applications and techniques of this data structure in the particular areas of digital imaging and remotely sensed data

IMAGE FUSION - FROM TIN TO PIXELS

Once all the low-resolution images have been matched to a sub-pixel level, TIN structures are generated for each image and the vertices (or pixels) are projected or mapped on a regularly spaced high-resolution grid (see Figure 3). In the idealized higher resolution arrangement of Figure 3 the 4 low-resolution images in Figure 4(a) are taken with sub-pixel shifts of half a pixel in the horizontal, vertical and diagonal directions in relative to the top-left image.

Their pixels can then be interleaved to generate an image with a magnification factor equal to 4, that is, Figure 3(b) contains 4 times more real pixels than any of the low-resolution images. However, in practice, these shifts are randomly distributed due to the irregular nature of the sub-pixel motions and the relative rotations of each triangulated frame in the sequence.

Therefore, these random motions or shifts must be known precisely in order to create a regular and refined grid of interpolated pixel values. Since the interpolation process is an estimation process which determines the pixel brightness which would exist on the intersections of a regular grid using randomly spaced pixel locations (representing the TIN vertices or pixels of the low-resolution images), several interpolators may be used depending on the application and accuracy requirements.



(a) (b)

Figure 3. An idelaised higher resolution grid. Triangulated vertices (pixels) of the low-resolution shifted images in (a) are mapped on a uniform grid (b), thus creating an image with four times more pixels than any of the images in (a)

One of the methods for interpolating scattered data to a uniform refined grid is referred to as Kriging [3]. This geo-statistical method was considered because it is a statistical interpolation technique that contemplates both the distance and the degree of variation between known data points when estimating values in unknown areas.

A kriged estimate is a weighted linear combination of the known sample values around the point to be estimated. The interpolation process determines the value of the points at grid intersections to have decimal figures. Hence, the values are rounded to the nearest integer which is the norm for representing pixels brightness within a raster image.

Kriging permits the user to classify weights that result in optimal and unbiased estimations. It attempts to minimize the error variance and set the mean of the prediction errors to zero so that there is no over- or under-estimates. An important feature of Kriging, as compared with other image or surface interpolators, is that it gives an estimate of the error at each interpolated point, thus providing a measure of confidence in the modelled surface (i.e. a digital image).

TESTS AND RESULTS

All tests were carried out on TIN models made from grey-scale images, as they usually better depict possible artifacts. However, colour images could have also been processed as they are images that can be considered having three different channels of grey-scale brightness levels.

The proposed enhancement process was tested for video imaging applications using simulated data. In this controlled and proof of concept experiment, the 'true' image was known prior to the enhancement and accordingly the accuracy of the enhancement and some of the factors that would influence the process could be examined and measured. The objective was to assess visually and statistically the performance of the proposed image enhancement technique.

A grey scale image referred to as luisa.tif (240x320), was scanned and stored with no compression applied to it. This image is shown in Figure 4(a). An image thus obtained was considered to be an image which would preserve the essential integrity of its grey-scale information. A total of 64 low-resolution images of luisa.tif (60x80) were also generated in the same way by setting the scanner resolution to be 4 times less than the original luisa.tif. These 64 low-resolution compressed images were scanned randomly and stored as JPEG with an unknown level of compression.

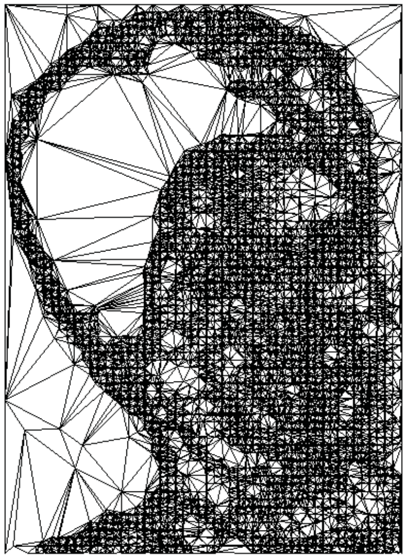
The main reason for using a conventional off-the-shelf scanner, rather than a standard digital video camera, was the flexibility that the scanner offered with respect to the selection of the image format and spatial resolution. Uncompressed and/or compressed images could be retrieved at various scanner resolutions and be stored in any desired image format.
Also, during the scanning process rotations were not applied to the images as they would have added further parameters to the enhancement process, thus diminishing the strength of the deductions reached in the tests. Correlation evidently exists amongst a video camera and the orientation parameters such as tilts, rotations and affinity/obliquity of the sensor, and, in a constrained test where the intention was to intrinsically prove the use of a process to enhance image resolution, it was assumed imprudent to add such constraints.

One of the 64 low-resolution compressed images is shown in Figure 4(b). The tolerance *d* for all the TIN models of the low-resolution images was selected as *d*=2 grey-scale brightness values. Figure 4(d) shows visually the results of fusing only 36 TIN representations of the 64 low-resolution images. Figure 4(c) depicts the fusion of all the TIN models of the 64 low-resolution images. While all 64 low-resolution images were used in the enhancement process, only 36 of the initial 64 contributed to improve the accuracy of the enhancement.

In addition, the number of TIN vertices processed to attain the result in Figure 4(d) was approximately 28% less than the total number of pixels included in the original luisa.tif. The enhancement process described in section 1 was applied to the 64 compressed images using a spatial enhancement factor (the ratio between low-resolution and high resolution grid size) of 4. The R.M.S.E. of the differences of the brightness values between the enhanced composite and the original luisa.tif was 3.6 with maximum and minimum values ranging between +5 and -4 grey-scale brightness values respectively.

(a)

(b)



(c)

(d)

Figure 4: The original high resolution image (240x320)) of luisa.tif (a) whereas (b) is one of the 64 down-sampled/compressed and shifted images used in the enhancement. (c) is the TIN representation of merging the 64 low-resolution images whereas (d) is the reconstructed enhanced composite derived from the TIN representation in (c).

The R.M.S.E. of the differences of the brightness values between the enhanced composite and the original luisa.tif was 3.6 with maximum and minimum values ranging between +5 and -4 grey-scale brightness values respectively.

By way of comparison, Table 1 gives an indication of the accuracy achieved using six other image reconstruction interpolators. The table was obtained by computing the R.M.S.E. of the differences of intensity values between the original image of the *cameraman* and the higher resolution composites computed using the interpolation techniques listed in the table. The lower R.M.S.E. value of *Universal Kriging* demonstrates the validity and the overall accuracy improvement of this interpolation technique.

Table 1. R.M.S.E.. of differences of pixels intensities between the original image of the cameraman and the higher resolution composites using various interpolation methods.

INTERPOLATION TECHNIQUE	**R.M.S.E. OF DIFFERENCE OF INTENSITY VALUES**
Nearest neighbour	19
Weighted average	14
Least squares plane	12
Linear interpolation	12
Bicubic interpolation	11
Papoulis-Gerchberg	10
Universal Kriging	7

CONCLUSIONS

In this paper, the problem of recovering a high-resolution image from a sequence of low-resolution and compressed images was considered. Low-resolution images contain aliased information. Accurate sub-pixel shifts within the low-resolution video sequence facilitate the recovery of spatial resolution from the aliased information. Notable findings encountered in this study are:

(a) The conversion of a set of low-resolution grey-scale images of the same scene into TIN structures which can be fused or combined so as to produce a composite raster image having more details than any of the individual low-resolution images.

(b) A method for enhancing the resolution of a compressed sequence of low-resolution triangulated images of a common scene using a Kriging interpolation approach has been presented where the matching or registration of the low-resolution images required for the enhancement relies on a matching method that can achieve fractional accuracies of +/-0.01 pixels. Comparison Kriging with present interpolation methods was also presented

(c) The application of the enhancement process has been demonstrated in 2D using controlled experiments which simulated a video sequence.

(d) The TIN structure approach to multi-frame image enhancement allows for redundant vertices (i.e. pixels) to be discriminated from the final enhancement process thus improving processing time while producing acceptable results.

More research is required to assess the accuracy of the enhancement process in the presence of rotations amongst the low-resolution input images and the presence of random noise. Refinements to the proposed method are being undertaken to increase the accuracy achievable for a variety of scenes of variable image entropy. Tests related to investigating the performance of a TIN approach to image enhancement whereby both sensor and object are dynamic, and the illumination is non-uniform, are presently being undertaken.

REFERENCES

[1] T. A. Slocum, R. B. McMaster, F. C. Kessler and H. Howard "Thematic Cartography and Geo-visualization". Pearson; 3rd edition. 576 pages, 2008.

[2] P. Vandewalle, S. Susstrunk and M. Vetterli. “A Frequency Domain Approach to Registration of Aliased Images with Application to SR”, EURASIP. Journal on Applied Signal Processing. 2005.

[3] B. I. Harman, H. Koseoglu and C. O. Yigit. "Performance evaluation of IDW, Kriging and multiquadric interpolation methods in producing noise mapping: A case study at the city of Isparta, Turkey". Elsevier. Volume 112, Pages 147–157. 2016.

[4] J. Kohout. "On Digital Image Representation by the Delaunay Triangulation". PSIVT 2007, December 17-19 pp. 826-840, 2007.

[5] S. Cheng, T. Dey and J.R. "Shewchuk. "Delauney Mesh Generation". Chapman and Hall/CRC Computer & Information Science Series. 410 pages. 2012.

[6] P. Milanfar "Superresolution Imaging" CRC Press. 490 pages. September 28, 2010.

[7] L. Simpkins and R.L. Stevenson, "An Introduction to Super-Resolution Imaging." *Mathematical Optics: Classical, Quantum, and Computational Methods*, Ed. V. Lakshminarayanan, M. Calvo, and T. Alieva. CRC Press, 2012. 539-564.

[8] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and Robust Multi-frame Super-resolution", IEEE Transactions on Image Processing, vol. 13, no. 10, pp. 1327–1344, October 2004.